Mapping the (R-)Evolution of Technological Fields

A Semantic Network Approach

Jurowetzki, Roman; Hain, Daniel S.

Published in:
Lecture Notes in Computer Science

DOI (link to publication from Publisher):
10.1007/978-3-319-13734-6_27

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

Mapping the (R-)Evolution of Technological Fields – A Semantic Network Approach ⋆

Roman Jurowetski and Daniel S. Hain

Aalborg University, Department of Business and Management, IKE,
Fibigerstrde 11, 9220 Aalborg, Denmark
{roman@business.aau.dk,dsh@business.aau.dk
http://www.ike.aau.dk

Abstract. This paper suggests a method to map the evolution of technological trajectories by using unstructured text data. Combining techniques from the fields of natural language processing and network analysis, we are able to identify technological fields as overlapping communities of knowledge fragments. Over time persistence of these fragments allows to observe how these fields evolve into trajectories, which may change, split, merge and finally disappear. As empirical example we use the broad area of *Technological Singularity*, a umbrella term for different technologies ranging from neuroscience to machine learning and bioengineering which are seen as main contributors to the development of artificial intelligence and human enhancement technologies. Using a socially enhanced search routine, we extract 1,398 documents for the years 2011-2013. While we can identify consistent technology fields in static document collections, more advanced ontology reconciliation is needed to be able to track a larger number of communities over time.

Keywords: Technology forecasting, natural language processing, network analysis, dynamic community detection

1 Introduction

Understanding the pattern and drivers of technological change is a crucial precondition to formulate meaningful long-term research and industry policy. This development usually happens along technological trajectories, within a *scientific paradigm* [1]. Apart from defining the boundaries, a paradigm also provides a set of generic technology artifacts that serve as interface, allowing for interaction and re-combination of knowledge and technologies between trajectories. In this paper we present a framework and methodology to understand, illustrate and analyze technological change and the (co-) evolution of technological trajectories using large amounts of unstructured text data from various sources on the internet.

⋆ We would like to thank Dan Mc Farland, Dan Jurafsky, Walter W. Powell, all participants of the ISS 2014 Jena conference, the KID 2014 Nice workshop, and the Stanford Network Forum for invaluable inspiration and feedback. All opinions, and errors, remain our own.

Many modern technologies are indeed characterized by the combination and re-combination of components from different trajectories. The rapid progress of ICT technology led to its penetration of virtually all areas of social and commercial activity, and the development of common data transfer protocols and interfaces is said to make heterogeneous technology components more compatible with each other. The continuation of these dynamics on a higher level of aggregation can be seen in the currently evolving Internet of Things (IoT). Here the connection of existing technological artifacts with sensors and communication devises is supposed to enable many new applications across formerly separated technological trajectories. The underlying theme has been widely discussed in literature on technology combination, diversity and innovation [2–4]. We argue that today - as briefly outlined in the examples above - we face an accelerating deterioration of the technological burdens to combination through growing complementary of components and modularization [5, 6]. In order to understand innovation activity in many technological fields, it thus becomes important to uncover the dynamics of these processes of recombination.

During the last decade we have witnessed tremendous growth of freely available digital information, often in the form of unstructured text data from sources such as web-sites and blogs, the written communication of entire communities in forums or via e-mail, and repositories (e.g. SSRN or recently Researchgate). The topicality and sheer amount of such data bear great opportunities for social science research. Yet, commonly used analytic approaches, are limited to in-depth case studies [7, 8], quantitative methods depending on data such as patents [9] or scientific publications [10], and more generic simulation models [11, 12]. While our understanding of the emergence and evolution of technologies has greatly benefited from such studies, they either require massive effort to qualitatively analyze such complex interaction patterns in technological space, or rely on quantitative data only available with non-negligible time delay, and only relevant for certain technology domains, often underestimating the context in which technology is used.

Our attempt is to provide a method to map the development of technologies by using large amounts of unstructured data from various sources by combining techniques from the fields of natural language processing and network analysis. Conceptualizing technological change as changing interaction patterns between technology fragments, and their clustering in space to technological fields, and in time to technological trajectories, we provide a framework as well as a methodology to deeper understand the co-evolution of technology. We use the case of *technological singularity* to illustrate our approach graphically as well as with key measures derived from network analysis.

The remainder of the paper is structured as follows. Section 2 reviews and discusses literature and concepts of technological change, and provides a theoretical framework for our approach to be developed in the following. Section 3 reviews empirical work analyzing technological change and discusses merits and drawback of applied methods. In Section 4 we provide a novel methodology how to map technological change and the evolution of technological fields, which we illustrate in Section 5 at the case of *Singularity* technologies. Finally, Section 6 concludes, provides implications for theory, empirical research, and suggests applications for science and industry policy.

2 Conceptualization of Technological Change

The development of technology is contextual to the development of industrial structures, or more broadly the different factors that act as focusing forces upon the direction of technological development [13]. It is also understood as happening within broader *technological paradigms* [1] and linked to a particular generic application. Within this framework technology is perceived as a mean to problem solving in a particular context. The same problem could easily be solved in various other ways using other technologies. *Technological trajectories* here represent pathways, which span across the multidimensional space defined by the paradigm [13]. They do so by generating technologies that respectively can contribute with alternative solutions to the general task that is defined by the paradigm. Overall that suggests competition between the trajectories. Yet, they can also be compatible to each other.

Many of the technologies developed since the publication of the framework are characterized by the combination and recombination of components stemming from different trajectories, perhaps even different paradigms. The rapid progress of ICT technology led to its penetration of virtually all areas of social and commercial activity. This dynamic paired with the development of common data transfer protocols and interfaces is likely to have made technologies from different trajectories more compatible with each other. In fact, [14] even assert that in the presence of network externalities, compatible designs make even sense among competitors, what would provide an additional incentive to support compatibility.

A recent and very obvious example for this development is the smartphone. The combination of voice and data communication with GPS, camera, compass and accelerometer technologies, bound together by a miniature touchscreen-computer opened up for a uncountable number of not anticipated applications. Various standardized wireless connection technologies like bluetooth or WiFi allow for compatibility with many other external devises, thus increasing the functionality and re-purposing the phone. Figure 1 illustrates this example. The continuation of these dynamics on a higher level of aggregation can be seen in the currently evolving Internet of Things (IoT). Here the pairing of everyday objects with sensors and communication devises is supposed to enable many new applications in a variety of contexts, and potentially triggering many disruptive innovations [15].

The underlying theme has been widely discussed in literature on invention and complex adaptive systems [3]. Drawing on work in theoretical biology [16], evolution is conceived as a process of recombination of novel and existing component technologies. The result of such a development, a technological system, can also be understood as a complex system with a number of elements that collectively fulfil a single or various goals [17]. The number of potential combinations in such a system is an exponential function of the number of elements. Yet, the amount of possible or useful combinations is moderated by the interdependencies or *epistatic relations* of the components. Interdependence is understood as a functional sensitivity of a system to changes in these constituent elements [3], meaning that a change in one element will affect the functioning of the particular element and the functioning of those that are epistatically related [18]. The level of complexity in such systems is thus determined by the number of elements and their respective functional interdependencies. Too high levels of interdependence might make possible combinations difficult to detect or costly to achieve, while a lack of interdependence implies no difference in functionality across configurations [3]. As an alternative to intermediate levels of interdependence, modularity has been discussed in the literature [5, 6, 19]. Modularization of systems aims at the development of standardized interfaces between more discrete elements to mediate interdependence [20], thus allowing to decrease the overall complexity while maintaining or even increasing the number of possible recombinations.

We argue that today, we are witnessing a rapid decline of the burdens to technology-combination through efficient modularization between components within artefacts such as the smartphone and on a higher level, where standardized interfaces opened up for combination. Embracing this line of thought, we present an framework and methodology geared towards the analysis of the evolution of such interdependent technology systems.

3 Measurement of Technological Change – State of the Art

Empirical research on technological change has a long tradition in different academic communities. Generally technology exists to fulfill or support some societal functions through direct application or indirectly through derived products, is thus always embedded in and framed by a societal, political and organizational context, which co-evolves with it [21]. Work by sociologists of science within the STS (Science, Technology and Society) tradition, has produced many concepts and valuable insights into processes of systemic technological change [22]. The work often relies on detailed description of the complex multidimensional setup around the studied technology and sheds light on the variety of factors that (can) influence and shape its development [23].

A substantial stream of more positivistic research in the fields of industrial economics and sciencometrics is based on patent data as an approximation for technological development. Research so far mostly incorporates patent data as aggregated numbers to explain differences in scale [24], or in a network representation to explain structural differences [9, 25] in the development of technologies across countries and industries. Patent data has also been used to study invention as a recombination process [3, 4, 26].[1] Alternatively, similar research also

[1] However, besides its merits and easy accessibility, there are widely recognized limits in the use of patent data [27, 28] such as the high variation of importance across

utilizes the assessment by industry experts to delimit and quantify development within and across technologies [29]. [30], suggest for instance the use of industry experts to delineate technological systems.

Most recently, social scientists have also started to deploy methods from the fields of computational linguistic and natural language processing to advance empirical research on the development of science, technology and other bodies of knowledge In their essence, such linguistically informed methods are capable of identifying patterns of language usage in large bodies of text and communication. They range from simple measures of (raw or somewhat weighted) word co-occurrence across documents, corpora and over time [31], to complex probabilistic language and topic identification models [32], which lately started to gain traction in the social science [33–36]. Such models basically identify larger topics by fitting a linguistically informed probability model which tries to predict them using text and meta information of the corpus under investigation. Such topics by nature are rather descriptive and aims to understand how language is used by a certain set of actors to describe and differentiate real-life phenomena. For instance, the interesting variety of lead-lag models which groups of actors, such as universities [37] or outlets [38] influence the formation of topics, and which adapt instead.

4 Analizing technology Evolution: Dynamic Semantic Network Approach

In this section, we provide a conceptual model to how to map the evolution of technological fields embedded in a larger technological system based on large amounts of text data. A summary of the method pipeline is illustrated in Figure 6

4.1 Conceptualization and Definition

The representation of systems of interacting elements as networks has brought fresh perspectives and insights to the analysis of complex phenomena from the biological to the social sciences [39]. As discussed earlier, technology can be framed as a system of interdependent components [40] within their respective trajectories of development [13].On an abstract level, one can imagine such a technological paradigm or technological system projected only in technology space as a system of interacting elements. On the lowest level of aggregation, we find what we call *technology fragments*, which are atomic, non-reducible repositories of scientific/technological knowledge needed to fulfill a certain and narrow task. In scientific, technological and industrial applications such as machines, software and other devices which we call *technological artifacts*, such fragments are linked in a functional relationship to produce some output. On a again higher level, sets of complementary and substitutional artifacts form a a *technological*

industries and countries, and over time and the long delay between the time research is conducted and the corresponding patent publication.

field. Over time, such field develop along Dosian technological trajectories, where accumulated sets of common configuration pattern partially reproduce over time and set the foundation for further combinations.Again, fragments and artifacts originating from one field might be reconfigured and redeployed in a different field to fulfill the same or even a different purpose. Furthermore, fragments as well as artifacts might not even mainly belong to one field, but be so generic in the nature of task they fulfill, that they can be deployed equally across multiple fields.

Such a conceptualization of technology evolution comes pretty close to how [41] describes the innovation process, as the recombination of existing resources in a novel way. It furthermore has the advantage that it allows us to envision technology evolution as a developing network, and deploy the rich toolkit of network analysis and visualization. In summary, our conceptualization of technological change and evolution, and the suggested methods to analyze it, is based on the following assumptions:

Assumption 1: Knowledge fragments are atomic, non-reducible repositories of scientific/technological knowledge

Assumption 2: Co-location of technology fragments in documents imply a functional relationship

Assumption 3: Technology fragments can be arbitrary combined and recombined to form functional technological artifacts

4.2 From unstructured Text to Technology Fragments

Having defined and extracted the relevant text documents that in the optimal case address a particular technology, it is necessary to reduce them to a *machine readable* representation. Typically, this takes the format of a bag of words (BOW), a line-up of thematically relevant keywords, usually nouns and bi-gram noun phrases. The key assumption of this type of NLP applications is that statistically significant co-occurrence patterns of concepts across the corpus is indicative for actual association between them. For our means, the goal is to reduce each document to the contained technological concepts. Instead of using a probabilistic approach that stepwise excludes text-elements that are definitely not a technology, we try to detect mentioned technologies in the data. This task falls into the category of named *entity extraction*, which typically relies on tagged dictionaries and string-matching rules to identify the required concepts. Entity recognition has recently also become important in the context of the so called *semantic web*. The aim of this development is to make more data on the internet machine-readable, structured and thus allow for more machine learning applications and cognitive computing in general. A number of applications related to this development target the identification of different concepts in unstructured text, among others technological and industrial terms. The advantage of these semantic web tools is that they are supported by large centralized, constantly updated and optimized dictionaries and *intelligent* disambiguation functions. The result of a successful entity extraction returns a collection of documents that only contain the mentioned technology terms and their document appearance frequency.

4.3 Network Creation

In a first step, one could create a network with the corpus documents as nodes, then vector space modeling and represent them as vectors defined by the respective combination of contained concepts. This representation allows to calculate pairwise similarities between the documents. The result is a fully connected weighed network with documents as nodes and corresponding similarities as edges. Now clustering or community detection algorithms can be used to identify technological fields, represented by document communities discussing them, as suggested by [42]. Yet this approach has two disadvantage: First, technology fragments are only indirectly represented in networks as node characteristics, what means that many powerful measures in network analysis (such as centrality, betweenness, etc.) are not directly available to describe them, but only the documents containing them. Second and related, nodes representing documents are not suitable for a dynamic analysis, since they are only associated with one observation period. Thus, one can either construct a cumulative network that only grows, or a network with a complete node turnover every period.

For that reasons we have chosen to *liberate* the terms from their *document boundaries* while maintaining the latent semantic similarity structure that is defined by their co-occurrence in documents. We construct a 2-mode network consisting of the distinct technology fragments as nodes of interest, and the corresponding documents as nodes in the second mode, linked by the pairwise cosine similarity between the vectors of the particular term and document within a 400-dimensional vector space as edges (see. fig. 3). The vector space is defined by training a LSI (Latent Semantic Indexing) model [43, 44] on the full corpus of documents, which is spanning along the whole timeline.[2] Rather then informing about the presence of a term in a given document, the weighted edge indicates to witch extent the term is semantically close to the entirety of the other terms contained by the document (see. Figure 2).

When separating such a network in time slices by selecting only the documents and the distinct terms that existed during a determined period, the node set of the first mode stays stable, since terms (as opposed to documents) tend to reappear over time. When projecting this network in the technology dimension, we end up with a one-mode network of technology fragments connected by the pairwise projected semantic similarity values, associated with the corresponding period.

Again, the underlying rationale is based on the assumption that co-occurrence in documents at least on an aggregated level also corresponds to a functional

[2] Before training the model, we apply TF-IDF weights to all terms within the documents. This appreciates the value of particularly important terms for the single document, while depreciating the value of generic terms that often occur across the corpus.

relationship between technology fragments. However, on a document level that will not always be true. While some documents will discuss separate and coherent technological fields, others might serve more as an overview on industry or research of a broader context, hence contain a collection of technological fragments from many otherwise distinct fields. Thus, we penalize documents containing more technology fragments in a similar spirit as the method used by [45], which can be represented by the following equation [46], where w_{ij} represents the edge-weight between nodes I and j, and p the corresponding documents. We end up with a node-set of technology fragments which might be imagined as a adjacency matrix with a stable composition over time. Figure 4 illustrates these nodeset properties in dynamic networks.

$$w_{ij} = \sum_{p} \frac{w_{i,p}}{N_p - 1} \tag{1}$$

4.4 Technological field detection

When analyzing the structure, function, and dynamics of networks, it is extremely useful to identify sets of related nodes, known as communities, clusters, or partitions [47]. We depict the evolution of broad technological paradigms as the change of structural properties in micro level interactions between atomic technology fragments. Changes in this structure reveal the ongoing emergence, functional combination and recombination of these fragments to assemble higher level technological artifacts. The atomic structure of technology fragments implies that they cannot resemble technological artifacts alone but only in a functional combination, and the systemic nature of technology that fragments will often maintain functional relationships to multiple other fragments. In the following we use state-of-the-art community detection techniques to identify communities of fragments characterized by dense internal interaction. Such communities resemble what we call a *technological field*.[3]

Overlapping Community Detection Early clustering and community detection algorithms, in network analysis and elsewhere, usually assumed that the membership of entities to one distinct groups. However, depending on the meaning of edges and nodes, many real life networks show a high overlap of communities, where nodes at the overlap are associated with multiple communities. This especially tends to happen when relationship of different quality are projected in a one-mode network [48]. Ones' social interaction network for instance may consist of family members, work colleagues, members of the same karate club

[3] An alternative approach would be to use to identify technological fields by the using topic modeling, an approach that lately started to gain traction in social science [36, 35, 32], create a two-mode network of terms and topics, and project it to an one-mode network of terms. However, for reasons described we here want to offer an alternative, where the topics are already identified using the powerful community detection methods offered by network analysis.

or other associations. The more diverse interests such a person has, in the more different communities this person will be assigned to. In the same way, the more generic the nature of a technology fragment or artifact, the more technological fields will have functional relationships with it. Some technological artifacts are that pervasive, they facilitate almost all other technologies in the way they work, such as by its time steam-power or nowadays semiconductors [49]. Most traditional community detection algorithms would in such a case detect communities somewhat resembling a core-periphery structure, with a central highly interconnected community surrounded by sparsely interconnected ones.

Since we want to avoid exactly that, we apply the link community detection algorithm proposed by [50], which is able to detect communities with highly pervasive overlap by clustering links between the nodes rather than the nodes themselves. Each node here inherits all memberships of its links and can thus belong to multiple, overlapping communities. By doing so, we owe respect to the overlapping and nested structure of technology, and are able to identify key technological fragments interacting with multiple distinct fields.

Dynamic Community Detection Former research within technology forecasting and technology field mapping, which relied on the analysis large text corpora in form of patent descriptions or scientific publication, usually identifies technological fields separately at different points of time [31]. Technological fields, however, do spontaneously reassemble themselves in a vacuum in whatever intervals. They emerge, mutate, grow or decline, split and eventually disappear. To owe respect the evolutionary nature of technology, we want to identify communities which are somewhat stable and thus to be found in multiple observation periods, but also allow technological fields to experience key-events in their life-cycle. Besides helping us linking changing communities over time, the identification of such effects in itself represent an interesting information.

We consider the following significant events a community might experience during its evolution, also illustrated in Figure 5:

- Birth & Death: The first time a community C_i^t is observed and not matched with an already existing community C_j^{t-1}. This community, however, does not have to be stable over time. We in fact expect a substantial share of communities to only appear in on period but not sustain. In this case, the birth is equal the death of the community, which occurs in the last period a community can be observed and in none of the following periods a match can be found.
- Pause: We relax the assumption that a dynamic community has to be observable in every period between its birth and death. Indeed, technological fields (which are the representation of the identified communities) might be more stable than the report-ing on them in articles, publications, tech-blogs or whatever might serve as corpus for our analysis. When allowing communities to pause for some periods but to be identified later again, we smoothen possible trend and hype effects in technology reporting.

- Merge: During the evolution of a technology it might happen that at one point two technological fields develop that much functional interdependence, that their main interaction with the rest of the system only happens between them, thus they merge and form a new technological field consisting of both of them. Technically that happens when our matching algorithm (explained in the following) matches two or more different communities with one dynamic community D_j in the previous period.
- Split: In the same manner, technological fields can also separate into independent disciplines. Technically a split occurs when one community C_i matches with two or more dynamic communities in the previous period.

Technically, we do so by applying a simple but effective heuristic threshold-based method allowing for many-to-many mappings between communities across different time steps proposed by [51]. Here we compare an identified community C_i^t in observation period t with the set of dynamic communities in the previous period $\{C_1^{t-1}, , C_J^{t-1}\}$ by employing the widely adapted Jac-card coefficient J_{ij}^t, calculated as follows:

$$J_{ij}^t = sim(C_i^t, C_j^{t-1}) = \frac{|C_i^t \cap C_j^{t-1}|}{|C_i^t \cup C_j^{t-1}|} \tag{2}$$

If the similarity exceeds the defined matching threshold $\theta \in [0, 1]$, both communities are added to the dynamic community D_i. Using this simple but efficient method has the advantage that is independent of (static) community detection in the observation periods, hence represents a somewhat modular approach. It can also handle overlapping as well as (with some minor adjustments) weighted communities. A major advantage of this approach is the separation of static and dynamic community detection is the high flexibility in the choice of suitable algorithms.

d

5 Demonstration Case

For the demonstration of our method, we intended to find an empirical case of technological development that would combine a large number of components from traditionally disconnected technological fields. Additionally, the *technology field* in focus should be yet in a formative stage and have a potentially strong and broad social impact to generate enough attention. The latter requirement is important as it is public interest that usually triggers high numbers of reporting and thus the production of text data, which this project builds upon. We decided to explore the field of *technological singularity.* Rather then a clearly delineated technological field, singularity represents a future scenario and an umbrella term that summarizes a number of developments in areas as diverse as neuroscience and 3D printing.

5.1 Empirical Setting: The *Singularity* Case

Technological Singularity as a term has gained momentum since the publication of Ray Kurzweil's book in 2005 [52]. Observing various measures of technological progress over time, he argues that most technologies improved their performance exponentially and therefore it is only a matter of a few decades until we will have reached a point in history when artificial intelligence will supersede human intelligence. The most powerful technological advancement of the 21^{th} century will happen when robotics, nanotechnology, genetic engineering and artificial intelligence reach a certain level of development and can be combined, what will potentially have disruptive consequences for society, culture and the human nature.

While many of the forecasts sound like science fiction, others seem plausible. Smartphones, for instance became a rapidly adopted human enhancement device and currently a number of different wearable technologies are entering the mainstream markets. Recently, *singularity* entered the European technology policy context, as a technological field within the Horizon 2020 programming. Since 2012, the Directorate General for Communications Networks, Content and Technology (DG CONNECT) is undertaking a foresight process to inform the ICT related programming of research to be financed under Horizon 2020, where *singularity* was identified as one of the 10 central technological fields. It is currently being examined closer to capture early signals and anticipate beneficial trends that should be supported within public research funding schemes.

5.2 Data Mining & Corpus Generation

Researchers, organizations and science journalists are increasingly using social media and the blogosphere to communicate findings and developments, far ahead of journal publication or conference proceedings. This makes microblogging platforms and in particular Twitter with over 200 million monthly active users (Feb. 2014) a valuable source of data. We now describe our data mining approach aiming at selecting relevant twitter updates by relevant users. Twitter's graph structure, built on followship links, is similar to citation networks in academic publications. This enables the construction of large directed graphs and allows applying network analysis methods, to identify central actors for a particular field or topic. For this study we constructed a large followship graph around the - somewhat arbitrarily selected - account *Singularity Hub*, which is an online news platform that actively reports on the topic. The initial *snowballed* network has 49,574 accounts. Using eigenvector centrality, we identify the most influential users and then manually reduce the number of nodes down to 34 twitter accounts. Figure 7 shows the most central fragment of the network. Coloring represents communities, detected by the Louvain algorithm and is mostly illustrative. Yet, we can see that the red cluster seems to contain all the central organisations that are present on twitter and focused on singularity and transhumanism like the H+ movement, KurzweilAI, David Orban and more. The green cluster is mostly populated with users that are related to robotics and the violet to software architecture. An overview of the selected user accounts can be found in Table 1.

Micro-blogged tweets (status updates) by these actors often contain links to research papers, popular media articles or blog entries that the selected user considers as worth communicating. For each of these accounts we extract up to 3,200 status updates starting with the most recent, 63k in total. We discard all updates that do not carry a link. Relevant tweets were then identified using a vector space model powered semantic search. The text content behind the embedded links - outside of Twitter is then extracted and processed, and finally represents our document corpus for further analysis.

5.3 Network generation and analysis

The documents in our corpus discuss technology from very different angles. Some talk about state-of-the art research in certain university labs, while others review the allocation of public research grants or venture capital investment strategies. When attempting to uncover functional relationships between technology fragments, it is crucial to avoid false positive caused by other relationships that are non-technical in nature, such as *being funded by the same investor*, or *developed in the same country.* As described above, we rely on entity extraction when condensing documents to BOW representations. In the particular case we use OpenCalais, a free web service that performs entity identification across 39 different concepts within submitted text data. The great advantage of *cloud-sourcing* in this case is given by the fact that the centralized machine learning algorithms of OpenCalais are trained on a very large amount of natural text and its dictionaries are constantly updated and optimized. An offline solution would hardly be able to compete in terms of performance and topicality.[4] In addition, OpenCalais provides ontology reconciliation and disambiguation.[5]

When inspecting the results we find clear technology terms such as *dna profiling*, *robotic surgical systems*, *clinical genomics* or *regenerative stem cell technologies*, which come fairly close to how we understand technology fragments. These terms narrowly describe technology deployed for a fairly delimited task. However, we also find boarder technology terms such as *stem cells genomics*, which span a somewhat larger field of applications and likely to include some of the aforementioned terms, and on an even more generic level terms such as *biotechnology* or *robot*.[6] While this clearly diverts from our theoretical framework, where we find on node level only functional interaction of atomic knowledge fragments, we do not consider that as worrisome for the analysis to come. Our main objective is to identify and delimit technological fields, what we will

[4] For an overview and performance evaluation of available systems see [53].

[5] Identified entities are in many cases enriched with metadata (e.g. profession for persons, ticker symbols for companies and geospatial coordinates for locations).Other detected entity types are not used in this analysis.

[6] In future iteration of this approach, a more conservative filtering of high-frequency contextual stopwords might decrease the presence of too general terms.

do mainly by clustering nodes applying an overlapping community detection algorithm. Most community detection algorithms assume communities to have a higher within than between connectedness. Our approach of choice is able to detect such *nested* communities without falsely treating them as one large community clustered around some broad and generic term. For a very first inspection of the nodeset we create a simple network of all documents connected by their similarity in terms of containing technology fragments, cluster them by applying the very common Louvain algorithm [54, c.f.], and plot them in Figure 8. For the three main communities detected we provide a tag-cloud, weighted by the fragments' TF-IDF scores. One can see at first glance that our *Singularity* corpus very broadly consists of three fields, where the biggest is centered around robotics, and the two others around (stem) cell and brain research, or to be more interpretative: Robotics, biotechnology and neuroscience.

As described before, we now construct a set of two-mode networks between this nodes and the documents in our corpus,[7], containing only documents published in the corresponding time period, which we choose to be half a year.[8] Finally, we project this structures on one-mode networks between technology fragments.

5.4 Identification of technological fields

Now we identify technological fields using the overlapping community detection approach proposed [50].[9] We first run the community detection separated for every time step independently. We do not a-priori set a fixed amount of communities, but rather set the cutoff at the point in the dendogram where the overall community density is optimized in every timestep.

Table 2 provides some statistics on the networks and communities, and their development. While subject to some fluctuation, the networks seem to develop from many to less nodes and edges, and to less but denser communities. This might indicate *Singularity* after an initial phase of experimentation to mature and establish more delimited fields and sub-disciplines.

Table 3 plots the network of knowledge fragments and their community membership for every timestep. Again, what can be seen is that *Singularity* appears to develop from a broad area without clear boundaries and high interconnectedness towards clearly delimited technological fields. However, we also find first hints that over time some very generic technologies such as *smartphones* and *artificial intelligence* appear to develop towards a very central position, where

[7] Vector space modeling is performed with the GENSIM package [55] within IPython, using LSI and a 400 dimensional model as suggested by [56]

[8] This choice has to be made according to the properties of the data to be analyzed, since best results can be achieved when the network structure shows some gradual change between the observation periods, but no radical turnover suggestion complete discontinuity. This corresponds roughly to a Jaccard index of the two networks somewhere between 0.2 and 0.8.

[9] We use the implementation of the link-community approach provided by [57] as package for the statistical environment R.

they serve as common interface between most other fields. While it seems unlikely that smartphones (as we understand them today) will be around for longer then a decade, their centrality in the singularity discussion can be understood as the importance of mobile devices that enhance our by nature limited interaction range. A more radical interpretation would be that the smartphone is already now making us to some extent transhuman. Artificial intelligence, on the other hand, is at the very core of the singularity debate.

Table 4 illustrates the composition of some selected communities.[10] The tag-cluster are a good way to visualize the interaction between the actual technologies, principal applications and challenges. The first cluster suggests for instance that an important area of application for biometric technologies in conjuncture with machine learning will be found within law enforcement. The second cluster addresses advancements in the area of augmented reality and connections to existent social network structures using primarily mobile devices.

6 Conclusion

The aim of this paper was to provide a novel method to map the development of technologies by using large amounts of unstructured data from various recent sources by combining techniques from the fields of natural language processing and network analysis. We identified 1398 relevant text documents all over the internet, using a social search routine that we built around the followship structure within the microblogging service twitter.

Using entity recognition tools from the semantic web area, we were able to reduce documents to technology-term representations and finally generate a semantic timestep network of technology fragments. Our community detection exercise identified many coherent technological fields within each community. Already the static clustering provides valuable insights in the emergence of new technological fields and applications for existing technologies. Overlapping community detection, allowed us also to identify certain *general* technologies that work as hubs between other technologies, stemming from a large number of different domains.

Yet, we find the results of the community-tracking over time unsatisfactory. The obstacle are *false negatives* that obstruct the identification of similar communities over time. While we, as humans, can see that very similar communities are present in successive timesteps, even though the contained terms are slightly different, the algorithm is unable to identify this because the terms are not identical. Our language is full of synonyms, metaphors and unregulated terminology. The reader of this article might get an idea that for us *clusters* and *communities* mean basically the same thing, but computers wouldn't stand a chance. While we are (yet) unable to *teach* the algorithm a deep understanding of ontology, we can try to normalize the terminology as far as possible. This future measure should increase the number of identical terms over time.

[10] For the sake of clarity, the technology fragments are weighted by their within-cluster centrality.

References

1. Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.
2. Koen Frenken, Paolo P Saviotti, and Michel Trommetter. Variety and niche creation in aircraft, helicopters, motorcycles and microcomputers. *Research Policy*, 28(5):469–488, 1999.
3. Lee Fleming and Olav Sorenson. Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30(7):1019–1039, 2001.
4. Lee Fleming and Olav Sorenson. Science as a map in technological search. *Strategic Management Journal*, 25(89):909–928, July 2004.
5. Carliss Young Baldwin and Kim B Clark. *Design Rules: The power of modularity*. MIT Press, 2000.
6. M A Schilling. Toward a general modular systems theory and its application to interfirm product modularity. *Academy of Management Review*, 2000.
7. Andrew Davies. Innovation in large technical systems: the case of telecommunications. *Industrial and Corporate Change*, 5(4):1143–1180, 1996.
8. Marko P Hekkert and Simona O Negro. Technological Forecasting & Social Change. *Technological Forecasting & Social Change*, 76(4):584–594, May 2009.
9. Bart Verspagen. Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10(01):93–115, 2007.
10. C.S. Wagner and L. Leydesdorff. Network structure, self-organization, and the growth of international collaboration in science. *Research policy*, 34(10):1608–1618, 2005.
11. Herbert Dawid. *Agent-based models of innovation and technological change*, volume 2, chapter 25, pages 1235–1272. Elsevier, 2006.
12. A Lopolito, P Morone, and R Taylor. Emerging innovation niches: An agent based model. *Research Policy*, 42(6):1225–1238, 2013.
13. Giovanni Dosi. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3):147–162, 1982.
14. Joseph Farrell and Garth Saloner. Standardisation, Compatibility, and Innovation, 2000.
15. Lieven Trappeniers, Mohamed Ali Feki, Fahim Kawsar, Mathieu Boussard, and Lieven Trappeniers. The Internet of Things: The Next Technological Revolution. *Computer*, 46(2):0024–25, 2013.
16. Stuart A Kauffman. *The Origins of Order*. Self-Organization and Selection in Evolution. Oxford University Press, April 1993.
17. Herbert A Simon. *The Sciences of the Artificial*. MIT Press, July 1969.
18. Koen Frenken. A fitness landscape approach to technological complexity, modularity, and vertical disintegration. *Structural Change and Economic Dynamics*, 17(3):288–305, September 2006.
19. S K Ethiraj and D Levinthal. Modularity and innovation in complex systems. *Management Science*, 2004.
20. R N Langlois. Modularity in technology and organization. *Journal of Economic Behavior & Organization*, 2002.
21. Sarah Kaplan and Mary Tripsas. Thinking about technology: Applying a cognitive lens to technical change. *Research Policy*, 37(5):790–805, June 2008.

22. Wiebe E Bijker, Thomas P Hughes, Trevor Pinch, and Deborah G Douglas. *The Social Construction of Technological Systems*. New Directions in the Sociology and History of Technology. MIT Press, May 2012.
23. W E Bijker. *Of Bicycles, Bakelites and Bulbs*. Toward a Theory of Sociotechnical Change. The MIT Press, 1997.
24. Keith Pavitt. R&d, patenting and innovative activities: a statistical exploration. *Research Policy*, 11(1):33–51, 1982.
25. Roberto Fontana, Alessandro Nuvolari, and Bart Verspagen. Mapping technological trajectories as patent citation networks. an application to data communication standards. *Economics of Innovation and New Technology*, 18(4):311–336, 2009.
26. Iwan von Wartburg, Thorsten Teichert, and Katja Rost. Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, 34(10):1591–1607, December 2005.
27. Zvi Griliches. Patent statistics as economic indicators: a survey. In *R&D and productivity: the econometric evidence*, pages 287–343. University of Chicago Press, 1998.
28. Keith Pavitt. Patent statistics as indicators of innovative activities: possibilities and problems. *Scientometrics*, 7(1):77–99, 1985.
29. Keith Pavitt. Sectoral patterns of technical change: towards a taxonomy and a theory. *Research policy*, 13(6):343–373, 1984.
30. Bo Carlsson, Staffan Jacobsson, Magnus Holmén, and Annika Rickne. Innovation systems: analytical and methodological issues. *Research Policy*, 31(2):233–245, 2002.
31. Ssu-Han Chen, Mu-Hsuan Huang, Dar-Zen Chen, and Siou-Zih Lin. Technological Forecasting & Social Change. *Technological Forecasting & Social Change*, 79(9):1705–1719, November 2012.
32. David Hall, Daniel Jurafsky, and Christopher D Manning. Studying the history of ideas using topic models. In *Proceedings of the conference on empirical methods in natural language processing*, pages 363–371. Association for Computational Linguistics, 2008.
33. John W Mohr and Petko Bogdanov. Introduction –topic models: What they are and why they matter. *Poetics*, 41(6):545–569, 2013.
34. Daniel Ramage, Evan Rosen, Jason Chuang, Christopher D Manning, and Daniel A McFarland. Topic modeling for the social sciences. In *NIPS 2009 Workshop on Applications for Topic Models: Text and Beyond*, volume 5, 2009.
35. Daniel A McFarland, Daniel Ramage, Jason Chuang, Jeffrey Heer, Christopher D Manning, and Daniel Jurafsky. Differentiating language usage through topic models. *Poetics*, 41(6):607–625, 2013.
36. Paul DiMaggio, Manish Nag, and David Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, 41(6):570–606, 2013.
37. Daniel Ramage, Christopher D Manning, and Daniel A McFarland. Which universities lead and lag? toward university rankings based on scholarly output. In *Proc. of NIPS Workshop on Computational Social Science and the Wisdom of the Crowds*. Citeseer, 2010.
38. Ramesh Nallapati, Xiaolin Shi, Daniel A McFarland, Jure Leskovec, and Daniel Jurafsky. Leadlag lda: Estimating topic specific leads and lags of information outlets. In *ICWSM*, 2011.
39. Mark Newman, Albert-László Barabási, and Duncan J Watts. *The Structure and Dynamics of Networks:*. Princeton University Press, 2006.

40. T P Hughes. The evolution of large technological systems. *The social construction of technological systems: New directions in the sociology and history of technology*, pages 51–82, 1987.
41. Joseph A. Schumpeter. *Capitalism, Socialism and Democracy.* Harper, New York, 1942.
42. Roman Jurowetzki. Exploring Technology Evolution in Danish Smart Grid Development - An NLP approach. In *DRUID Academy Conderence*, pages 1–25, January 2014.
43. Scott Deerwester. Improving Information Retrieval with Latent Semantic Indexing. *In Proceedings of the 51st ASIS Annual Meeting (ASIS 8́8), Vol. 25 (October 1988)*, 25, 1988.
44. Scott C Deerwester, Susan T Dumais, Thomas K Landauer, George W Furnas, and Richard A Harshman. Indexing by latent semantic analysis. *JASIS*, 41(6):391–407, 1990.
45. Mark Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review*, 64(1):016132, 2001.
46. Tore Opsahl. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 35(2):159–167, 2013.
47. Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663, 2004.
48. Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *arXiv preprint arXiv:1309.7233*, 2013.
49. Carlota Perez. Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, page bep051, 2009.
50. Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, 2010.
51. Derek Greene, Dónal Doyle, and Pádraig Cunningham. Tracking the evolution of communities in dynamic social networks. In *Proc. International Conference on Advances in Social Networks Analysis and Mining (ASONAM'10)*, 2010.
52. Ray Kurzweil. *The Singularity Is Near.* When Humans Transcend Biology. Penguin, September 2005.
53. Giuseppe Rizzo and Raphaël Troncy. NERD: evaluating named entity recognition tools in the web of data. In *Workshop on Web Scale Knowledge Extraction (WEKEX'11)*, 2011.
54. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
55. Radim Řehůřek and Petr Sojka. Gensim—Statistical Semantics in Python. 2011.
56. Roger B Bradford. An empirical study of required dimensionality for large-scale latent semantic indexing applications. In *Proceeding of the 17th ACM conference*, page 153, New York, New York, USA, 2008. ACM Press.
57. Alex T Kalinka and Pavel Tomancak. linkcomm: an r package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics*, 27(14), 2011.

Appendix

Fig. 1. Illustrative combination of technology components from different trajectories

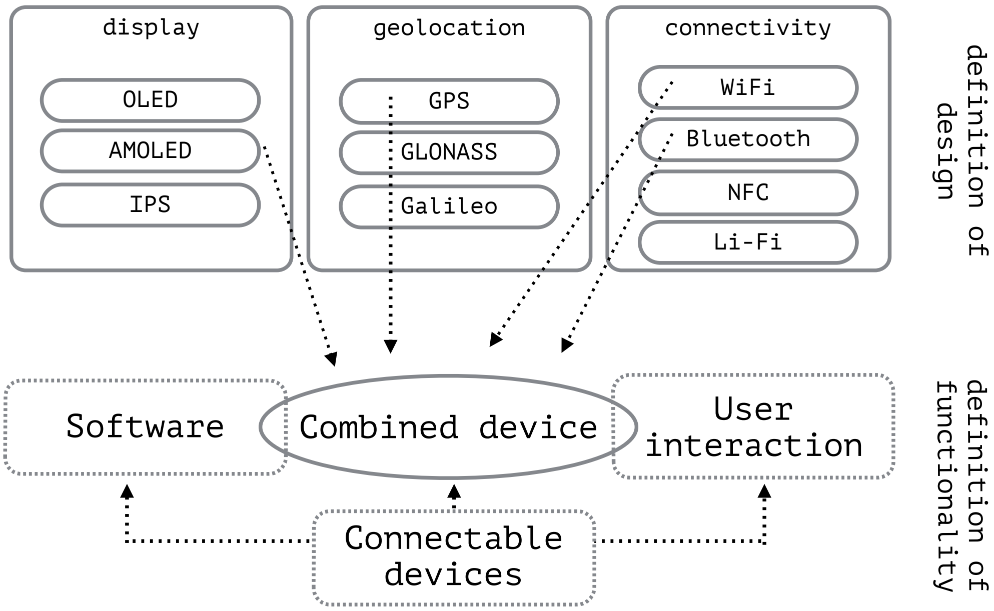

Fig. 2. Example of pairwise semantic similarity between terms and documents

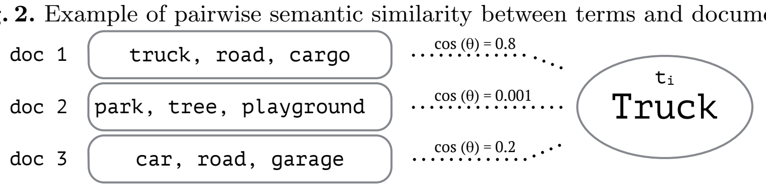

Fig. 3. Vector projection of terms and documents

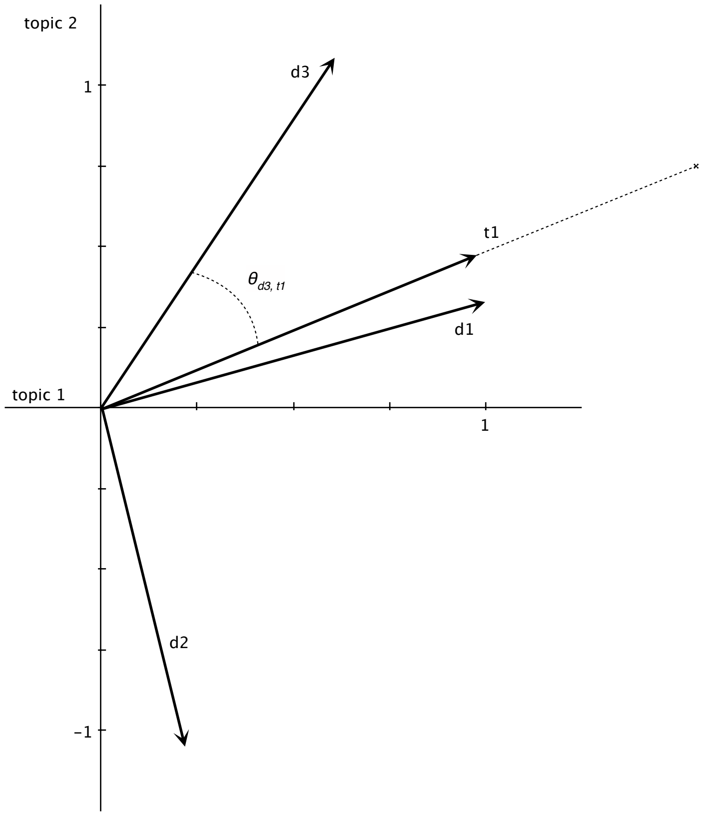

Fig. 4. Illustration of the development of a nodeset over time

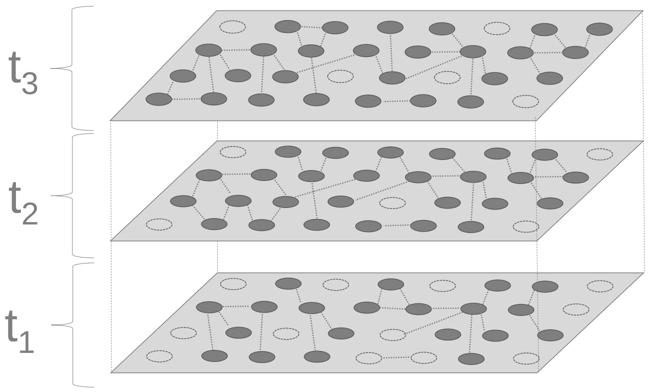

Fig. 5. Illustration of significant events in the evolution of communities

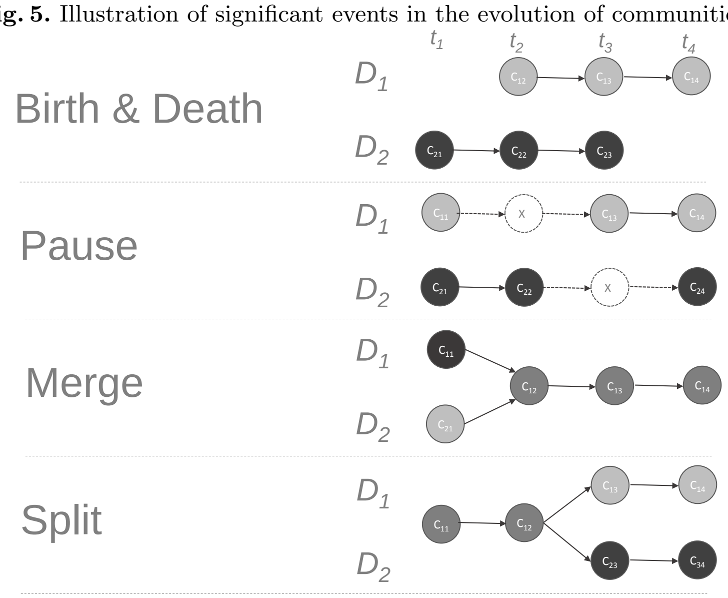

Fig. 6. Illustration of the method pipeline

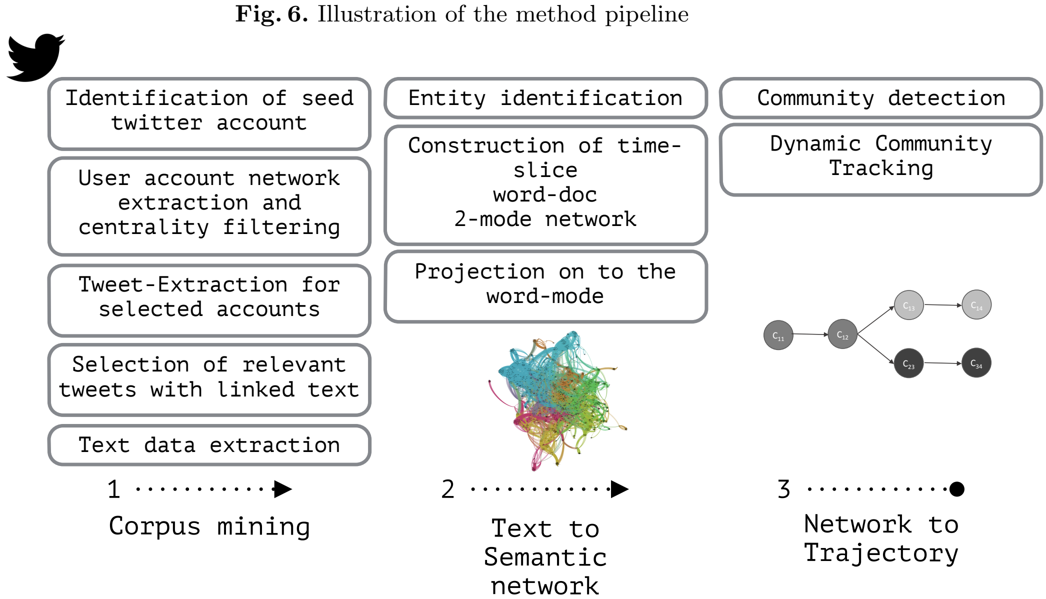

Fig. 7. The central fragment of the twitter account network with the finally selected profiles for text-extraction

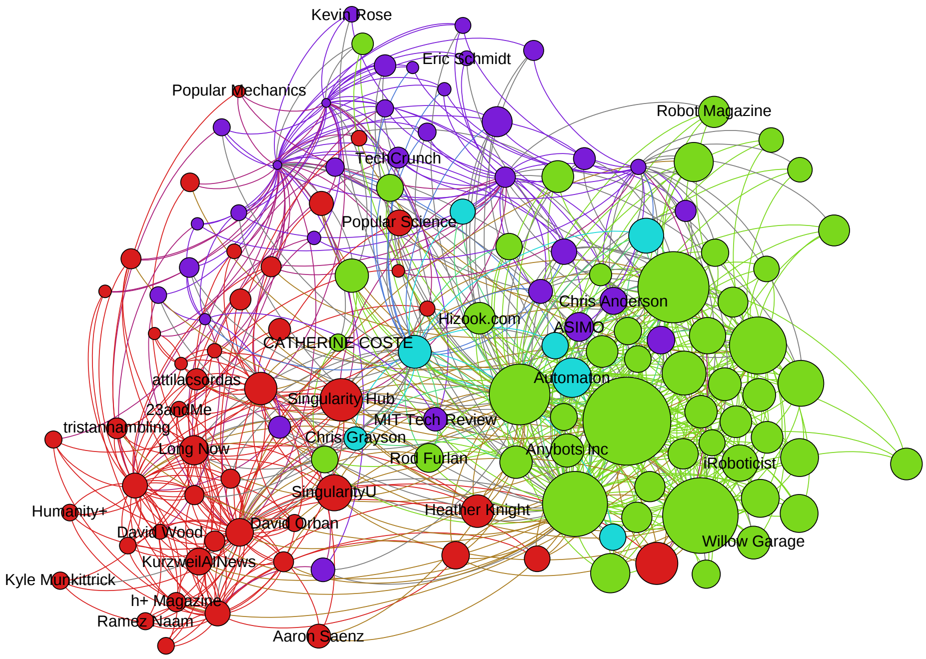

Table 1. Overview over the "expert" Twitter-accounts that were used for the text extraction

Twitter-id	Name	Location	Description
121684992	CATHERINE COSTE	Genomic Entertainment	MIT certificate in Genomics. Genomic & Precision Medicine UCSF. Blog: Ethics, Health & Death 2.0 - DTC Genomics
16870421	SingularityU	NASA Moffett Field, CA	Silicon Valley's leading experts on exponential technology. Follow @singularityhub @singularitylabs @suglobal @exponentialmed
6044272	Ramez Naam	Seattle	Author: Nexus / Crux / More Than Human / The Infinite Resource. Formerly a computer scientist at Microsoft. Interested in everything.
18705065	Humanity+	Global	Humanity+ is dedicated to promoting understanding, interest and participation in fields of emerging innovation that can radically benefit the human condition.
95661007	Kyle Munkittrick	Denver, NYC, San Fran	Bioethics: the unholy union of science, medicine, and philosophy. Blame no one but myself for what you find here.
16352993	Heather Knight		CMU Roboticist with a soft spot for interactive art & live robot performance: Founder @MarilynMonrobot, Director @robotfilmfest, Robo-Tech @RobotCombatSyFy!
28132585	Aaron Saenz		Writer for Singularity Hub, former Physics dude, Improv Comedian, Nomad
2443051	attilacsordas	Cambridge, UK	bioinformatician, EBI, regular Hadoop & R tinkerer, personal proteomics instigator, ex mitochondrial-stem cell biologist driven by healthy lifespan extension
15249166	Singularity Hub	NASA Moffett Field, CA	News network covering science, technology & the future of humanity. Follow @singularityu — Become HUB Member: http://t.co/wXGCjvICbk
16838443	KurzweilAINews	California/Mass	KurzweilAI (http://t.co/KD0Hq6D66p) is a newsletter/blog covering nano-bio-info-cogno-cosmic breakthroughs in accelerating intelligence
7445642	Chris Grayson	New York City / San Francisco	#Wearables / Advisor: http://t.co/ks23gJj30J / Prior ECD: http://t.co/Q4pSuEbFs4 / Events: http://t.co/bqLtGO3G5o & http://t.co/GrmHeG38Ir
16934772	tristanhambling	New Zealand	Tracking future, tech, nano, bio, neuro, info stuff, and anything new that scans past my event horizon. http://t.co/7aJFwAjkv7 also @futureseek
19004791	David Wood		Chair of London Futurists. Writer & consultant. PDA/smartphone pioneer. Symbian co-founder. Formerly at Psion and Accenture. Collaborative Transhumanist
15410587	Rod Furlan	Vancouver, BC	Artificial intelligence researcher, quant, Singularity University alum, Google Glass Explorer, serial autodidact, science lover & soon-to-be-robot
23115743	h+ Magazine	USA	h+ Magazine covers technological, scientific, and cultural trends that are changing human beings in fundamental ways.
743913	David Orban	New York, NY	CEO, Dotsub / Advisor & Faculty, Singularity University Analyzing and applying cycles of accelerating technological change. Flowing in wonderment.
19748200	Gizmag		I am a website about emerging technologies.
19722699	Popular Science	New York	Science and technology news from the future! Tweets from @RosePastore
138222776	Neurotechfuture	Boston, MA, USA	The future of life, humanity, and intelligence rests in the minds and hands of the innovators who envision, guide, and build it.
594718367	Grishin Robotics	New York	Everything about consumer robotics, connected devices & IoT. Published by the first robotics investment company. Founder - @dgrishin, feed editor - @Valery_Ka.
86626845	Eric Topol	La Jolla, CA	Cardiologist, geneticist, digital medicine aficionado, Editor-in-Chief, Medscape, author of The Creative Destruction of Medicine
15808647	MIT Tech Review	Cambridge, MA	We identify important new technologies deciphering their practical impact and revealing how they will change our lives.
101775759	Hizook.com	San Jose, USA	Robotics News for Academics & Professionals by Travis Deyle
44910688	Robot Magazine	Ridgefield, CT USA	The latest in hobby, science and consumer robotics.
16695266	ChiefRobot	Boston	Your daily dose of robots.
151648741	RobotoWear		Clothing for humans, inspired by robots. Robot t-shirts, hats, polos and hoodies.
87468736	Eric Tatro	Chicago, IL	Tweets about transhumanism, the singularity, AI, nanotech, biotech, robotics, life extension and human enhancement. All tweets and opinions are my own.
103516873	Willow Garage	Menlo Park, CA	Helping to revolutionize the world of personal robotics
6778032	Robert Oschler	Idaho	Artificial Intelligence and smart phone developer, currently focusing on speech recognition and natural language understanding applications and robotics.
18066713	robots_forever	Tokyo, Japan	Robot news, robotics research, combat and humanoid robot events, and other robot coverage from Japan.
8125922	Alexander Kruel	Germany	Transhumanist, atheist, vegetarian interested in math, programming, science fiction, science, language, philosophy, consciousness, the nature of reality...
22910080	Rob Spence Eyeborg	Toronto, Canada	We've built a wireless video camera eye. Tweets about privacy, cyborgs, prosthetics, eyepatches, Star Trek, The Bionic Man, and Augmented Reality.
7796912	Transhumanists	New York, NY	Singularity, Transhumanism, Artificial Intelligence, Human Enhancement, Stem Cells, Nanotechnology, Renewable Energy
15784353	Sensium		Revolutionary body monitoring for healthcare: wireless, intelligent, continuous, low-cost.
23116280	Popular Mechanics	New York City	The best in tech, science, aerospace, DIY and auto news. Customer Service: http://t.co/rYWTFWzg2R

Notes: Data extracted using the Twitter API in May 2014. Accounts can be freely accessed using https://twitter.com/intent/user?user_id=[insert here the twitter id]

Table 2. Network and community statistics over time

	2011, 2nd	2012, 1st	2012, 2nd	2013, 1st	2013, 2nd
N nodes	320	293	341	163	233
N edges	3,979	2,579	3,445	1,105	1,752
N communities	74	49	66	30	36
Max. community density	0.58	0.77	0.63	0.75	0.71
Max. nodes community	54	34	28	21	26

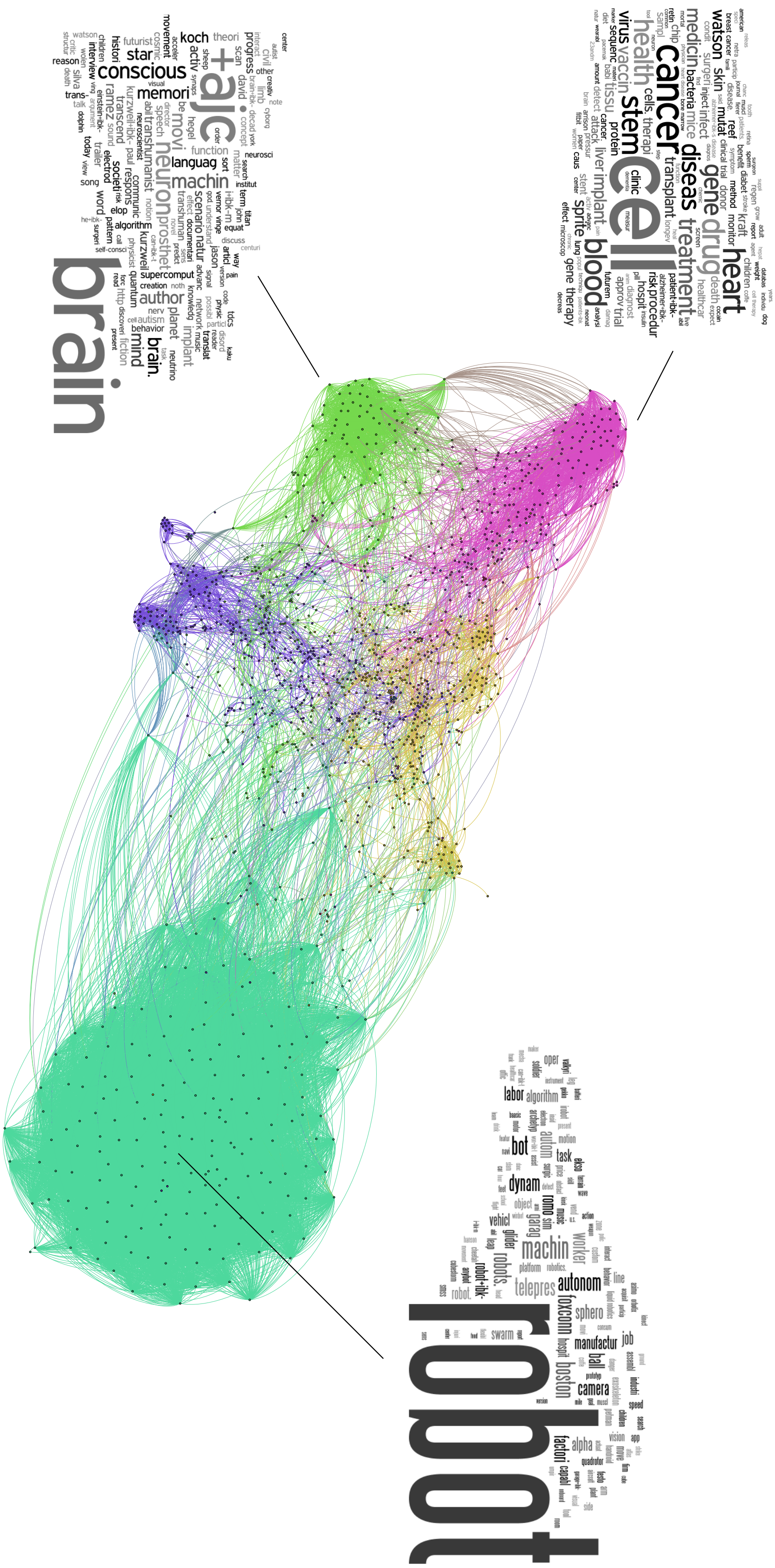

Fig. 8. Static Community Detection: Document similarity network of the whole corpus

Table 3. Network of Knowledge Fragments per Period after Overlapping Community Detection

2012, 1st

2012, 2nd

2013, 1st

2013, 2nd

Nodes are aligned according to their main community, represented by the number outside the circle. Node size is scaled by number of communities the node belongs to. Multi-community membership is also indicated by multiple node color

Table 4. Exemplary identified technological fields and their knowledge fragments

Biometrics & Law Enforcement

Ubiquity & Social Networks

Genomics

3D Printing

Nodes term representing the name of the technology fragment represented as tag-cloud. Size weighted by the nodes within community degree centrality.